

UNITED STATES PATENT APPLICATION

OF

GEORGE HAMILTON HARVEY

AND

SUZANNE ELIZABETH HARVEY

FOR

**SYSTEM AND METHOD FOR STORING DATA USING A MACHINE READABLE
VOCABULARY**

SYSTEM AND METHOD FOR STORING DATA USING A MACHINE READABLE
VOCABULARY

FIELD OF THE INVENTION

- [1] The present invention relates generally to the storing and processing of non-numerical data in a machine-readable and machine-operable form. More specifically, the invention includes a process and system for storing and processing a vocabulary that represents all concepts in a form in which the meaning of each word is processed and stored by machine.

[2]

SUMMARY OF THE INVENTION

- [1] The invention provides a system and method for storing and processing words of a vocabulary that is structured to represent all concepts in a manner that the words are easily stored and processed by machine. The words are divided into a number of fields, each field having meaning with respect to the meaning of the word. The fields are stored and processed in a manner that allows the meaning of each field to be recognized by machine. The meanings of each field are processed to interpret the meaning of each word. This vocabulary of words as stored and processed by machine is particularly useful in fields such as artificial intelligence, natural language processing, and database processing.

- [2] Each word includes a number of word roots selected from a set of word roots. Each word root is in turn divided into fields, organized from the most to least significant in a manner that imposes a tree-type taxonomy on the word roots. Each field in a word represents a characteristic of that word. The most significant field provides a class for the word root. Successively less significant fields, as they exist, divide the word root into

successively less significant subclasses, each a more definite subset of the more significant subclass being divided. The least significant field provides the category, which is a subset of a next more significant subclass. The category is the most finely definite definition normally available in the set of word roots. Each class, subclass, and category has a value unique within its level of definition. A field within the root represents each level of the root taxonomy. The value of the field represents a part of the meaning of the root. The universal set of all concepts is divided into as many subsets as provided at the finest level of division

- [3] The roots are combined to form words. Each root combined to form a word represents a particular characteristic of the word. Together the meanings of the roots give particular meaning to the word. The roots representing the words each include similar fields representing similar levels of the tree-type taxonomy. Accordingly, all roots can be processed in a similar manner and can be processed in parallel.

- [4] The most universal of all concepts is taken as existence. For this reason, all classes are taken as subsets of existence. The first subset of existence is existence itself, as distinct from the other subsets, which are particularities. Similarly, the division of each class produces one subclass which has the same name as the class itself, along with other more particular subsets. This extends to the categories, and so the first category in each subclass has the same name as the subclass itself. Thus, the first category is "Existence" possibly a subset of the subclass of "Existence," and certainly a subset of the class of "Existence."

- [5] A word may also contain a field or fields that are not word roots. For example, a field may be composed of bits, each bit indicating the negation of a specific root of the word.
- [6] The vocabulary with these properties is versatile in that it enables all concepts to be represented by a series of fields that are easily stored and processed by computer. Each of the fields provides meaning to the concept and can be processed and manipulated to provide the meaning of the concept. The meanings of each root of a word are commonly independent of one another and thus may be processed independently. This independent processing of roots allows for fast processing as well as for subtlety in the definition of the word.
- [7] The above properties make the vocabulary particularly useful for machine storage and processing. Each word is easily represented in the number of bits contained by a processor register. A computer programmed to recognize the meaning of words presented in this form is capable of quickly determining the meaning of the word and can determine various nuances in the manner that the roots are combined. The computer can store concepts using this vocabulary that are directly related to the physical world, but independent of existing human languages. For the computer to work, however, a complete taxonomy is unnecessary. The computer can be provided with particular roots at particular levels of definiteness as required by the task the computer is to perform. The computer may also be provided the meaning for particular roots as the computer encounters new roots or determines a need to employ a new root. This versatile vocabulary allows the computer to efficiently process ideas through association.

BRIEF DESCRIPTION OF THE DRAWINGS

- [1] Figure 1 shows a word of the vocabulary of the present invention.
- [2] Figure 2 shows a root taxonomy of the present invention.
- [3] Figure 3 shows illustrative words at the root level as represented by the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

- [1] In the preferred embodiment of the present invention, data is stored in words by a computer. The composition of the words is particularly designed to allow the computer to process and store the words based on the real meaning of the word. The words each represent a concept. The words are represented in digital form as they are intended to be machine read, processed and stored. In the preferred embodiment, a word is represented by a number of bits equal to the number of bits contained in the processor register of a computer used to process the words of the vocabulary.
- [2] An example of a word of the present invention is shown in Figure 1. The word 10 is 64 bits long and is thus designed for particular use with a 64-bit processor. The word includes a number of roots 20. The roots are selected from a set that defines a taxonomy in which the roots have a one-for-one relationship with the bit-structure. The root is divided into fields 30 with each field representing a level of a tree-type taxonomy.
- [3] The taxonomy used to define roots is shown in Figure 2. In the taxonomy of the present invention, the most universal concept (taken to be existence) is divided at a highest level into classes. The taxonomy has a tree-type structure that is similar to the tree-type classification system originally used in Roget's Thesaurus. The taxonomy includes a number of levels of significance. In the example, the taxonomy includes three

levels: classes, subclasses, and categories. In the example, the taxonomy includes eight classes represented by a field of three bits in a root. These classes are general abstract subsets of the most general concept of existence. Each class is divided into further subsets of subclasses. In the illustrated example, each class is divided into four subclasses. Each subclass is divided into further subsets of categories. In the illustrated example, each class is divided into eight subclasses.

- [4] The illustrated taxonomy departs from Roget's system in that the number of branches from one level down to the next is fixed. Each class is divided into four subclasses. Each subclass is divided into eight categories. All concepts fall within a category. To ensure that each concept will fall into a subset at each level, each division includes a broad subset that is similar to the subset of the higher level. For instance as existence is the most universal of concepts, those concepts that do not fall within another class are classified in the class of existence. Under the existence class there is an existence subclass that encompasses all concepts in the existence class that do not fall within the relation, order and quantity subclasses. In a similar manner, there is a life subclass in the life class, a life category in the life subclass, and a human category in the human subclass. Under this system, every concept is assigned a class, a subclass, and a category.

- [5] Each root includes one field corresponding to each level of the taxonomy. In the illustrated example, a most significant field of three bits represents the class of the root. A two-bit field represents the subclass. A least significant field of three bits represents the category of the root. Each root in the example is thus represented in eight bits. These three fields are common to each root. The value of each field is directly related to the

meaning of the root. In the example, all roots having a value of three in the most significant field are concepts within the life class. Likewise, all of these concepts with a value of one in the subclass are concepts within the human subclass.

[6] The taxonomy described above may be altered in various manners consistent with the present invention. More or less than the three levels of the tree (class, subclass, and category) may be used. Each level of the taxonomy may include a greater or smaller numbers of subsets. However in keeping with the one-to-one relationship between the roots and the bit structure, each subset at one level of the taxonomy is divided by a power of two into the subsets of the next lowest level.

[7] A finite number of roots is defined using this tree-type taxonomy. Roots are combined to define words. Each word includes a certain number of roots. In the illustrated example each word includes five roots. Each root represents a characteristic of the word and is assigned using a defined algorithm. In the illustrated example, the first root, designated the base (BASE), represents the context of the word from the taxonomy described above. The base is the contextual essence of the word. In practice the base root may be determined by looking the word up in a reference of the taxonomy similar to Roget's Thesaurus and finding the class, subclass, and category of the word. The second root is designated the alternate (ALT). The second root supplements the base with another basic component of the word. In practice the alternate root may be determined by looking a word up in dictionary and finding the primary word of the definition and looking that word up in a reference of the taxonomy.

[8] The remaining secondary roots define other characteristics. One root represents the source or cause of the word (SRC). One root represents the destination or purpose of

the word (DST). The remaining root represents the mode or method of the word (MOD).

Each root provides additional meaning to the word in a subtle and nuanced manner that cannot be accomplished solely by employing the tree-type taxonomy. The tree-type taxonomy provides the basic connection between the field values and specific meanings. However, for each category of the taxonomy to have significant meaning, the meanings are relatively broad. By combining roots in this multidimensional manner, each value for each field has significant meaning. Each root narrows the meaning of each word, yet each root may be processed in a similar manner and in parallel to extract the meaning of the word.

[9] In Figure 3, a number of illustrative words are shown. As an example, an “electronics teacher” may be represented by an alternate root of teaching, a base root of worker, a destination or purpose root of electronics, a mode root of communication and a source root of knowledge as shown in Figure 3. The representations for a number of other words are shown in Figure 3. It should be noted that merely by looking at the most significant field of the alternate root it can be determined that all the words but “instructional experience” are related, as being in the life class. It is can also be determined that by looking at the entire alternate root that the first six words are more closely related, as being in the worker category. Each field can provide meaning to the word. If the field requires no meaning, the value of most general concept “existence” is used in the field. A computer can process and store each word based on the meaning provided by any field or any combination of fields.

[10] For words that are particularly susceptible to classification, certain roots may be conventionalized. For example, “cat” may be represented by ALT root of animal. Using

process discussed above, the base root would also be animal. As this combination provides little information and would be similar for all animals, the base root may be conventionalized. By convention, invertebrates are assigned the class value normally indicating space. Vertebrates are assigned class value normally indicating physics. Cold-blooded vertebrates will be assigned the subclass value normally indicating geography, while warm-blooded vertebrates are assigned the subclass value indicating weight. Fish are assigned the category value of lake. Amphibians are assigned the category value of marsh. Reptiles are assigned the category value of land. Birds are assigned the category value of rarity. And mammals are assigned the category value of density. The conventionalized roots are useful where the computer can easily determine meaning from the field values under the convention. These conventions are thus used where the word is better defined by further classification rather than by the standard characteristics represented by the roots. The conventions are also chosen in concert with the taxonomy so that the taxonomy may continue to provide some relationships. For instance, fish are assigned the category value of lake, while amphibians are assigned the category value of marsh. The conventions must conform to the tree-type taxonomy structure. The conventions merely indicate altered meanings of the values of the fields of the roots. The conventions thus use the class, subclass, and category fields that make up each root in a modified manner.

- [11] In addition to conventionalizing some roots, the secondary roots may define alternative characteristics for some alternate values or some alternate and base combinations. In this example, the secondary roots define alternative characteristics when the alternate root has the value that indicates animal. The source root indicates where the

animal lives. The mode root indicates what the animal eats. The destination root indicates the value of the animal to humans. In the example, “cat” has a source root value indicating land, and mode root value indicating animal, and a destination value of associate (which by convention is used to indicate pet).

- [12] Each word is comprised of roots that provide meaning to the words. The words may also include other indicators that supplement or alter the meaning of the roots. In the example of Figure 1, the word includes 64 bits. The word in the example also includes six negation bits 40. These indicators are used to designate whether each root should be negated or interpreted with an opposite meaning. Other bits in the word are used by convention where required. In the illustrated example, “cat” has five roots. The alternate value indicates that it is an animal and that the base root is conventionalized and the secondary roots have alternate meaning. The base value indicates a mammal. The secondary roots indicate that it is a land dwelling, carnivorous pet. In the case the five roots do not distinguish between a cat and a dog. By convention three further bits are used to indicate the type of carnivorous pet. Values of zero for general (unknown or other), one for cat, two for dog, etc. are assigned. The remaining eight bits may be used to further define the word where necessary. In this example, other bits could be used to indicate the weight or the breed of the cat.

- [13] The word may also include a connotative root 50 that indicates not further meaning of the word, but rather how the word is used. This connotative root 50 provides nuance of usage and indicates when the word is appropriate to use. The connotative root 50 does not have usefulness in the absence of human language. To give the computer an ability to understand such things as humor, anger, and attempts to be polite or insulting,

the values of the connotative root are used. The connotative value indicates which human language word should be selected when human language words have similar meaning. For instance, certain values of the connotative root will indicate whether a word is slang, vulgar, formal, or technical. Thus, given the concept of a burp, the computer is able to select between the English choices of "burp", "belch", or "eructation". With proper connotative values considered in the translation to English, "burp" is used in polite company, "belch" in crude usage, and "eructation" for medical usage. Similarly, an operator may tell the computer, "You have an obsolete processor and faulty memory" or the operator may say, "You are a dolt and a bubble-brain." Through the application of the connotative root, the computer is able to discern the insult in the later statement, but fail to see it in the former.

- [14] The vocabulary of this invention is easily processed and stored by computer. As previously described the computer should include a processor having a register for receiving the number of bits forming each word. Words matched to the processor in this manner are completely manipulated in a single cycle resulting in efficient processing. The words may also be transferred from memory or other storage media over data busses that transmit an entire word in one cycle. The vocabulary is formed of words represented in digital form and having a length chosen as the number of bits in the register of the processor. Each word has a similar form. In the preferred embodiment the words each include five eight-bit definition roots, an eight bit connotative root, six one bit negation indicators, and a further ten bits used for other indicators. To process these words the computer uses a relatively simple algorithm. In the preferred embodiment, the computer receives all of the bits of a word in a register of a processor.

[15] The processor is programmed to recognize the bits that comprise each field of the word. The computer initially processes the alternate root. Each root is processed in a similar manner. The value of the most significant field is determined, thus giving broad meaning to the root. The values of the other fields are determined down to the least significant field. Each value provides a narrower meaning of the root. The computer is programmed with the root taxonomy necessary to recognize the meaning of each field value. The computer is also programmed with algorithms to recognize the conventions applied to any words that the computer will use. However, regardless of any convention, each root contains the same fields. The computer may identify the field of each root using the same process. The computer is programmed to initially determine the meaning of the alternate root. The meaning of the base root is determined taking into account any conventions based on the alternate root. The meaning of the secondary roots are determined taking into account any conventions or alternate definitions based on the base and alternate roots. The computer then recognizes any adjustments or supplements to the meaning based upon the additional indicators.

[16] The computer may form and store words by using a similar algorithm. The computer in order to form and store a word first determines the alternate root. The basic component of the definition of the word is looked up in the root taxonomy to determine the alternate root. The field values for the class, subclass, and category of the alternate root are determined. The values of the remaining fields are determined by referencing the standard root taxonomy unless the alternate root indicates conventionalized values. The base root is determined in a similar manner based upon the basic context of the word from the root taxonomy. The other roots values are determined based upon characteristic of the

root as found in the root taxonomy. The computer is programmed to look to the specific conventions in lieu of the standard root taxonomy based on the value of the alternate root or the base root where appropriate. The field values are selected from the root taxonomy to describe the characteristic of the word define by the root. The process is altered to use specific conventions or to define alternative characteristics based upon the base and alternate root values. The resulting word is digital information that is the computer is able to process and store by conventional methods. The computer can cause words of this invention to be stored in conventional readable media including electronic media such as memory or magnetic media such as disks and tape.

- [17] Other embodiments, uses and advantages of the present invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed. The specification and examples are exemplary. The scope of the invention is set forth by the following claims.